[AI-PHI] 6ᵀᴴ SESSION
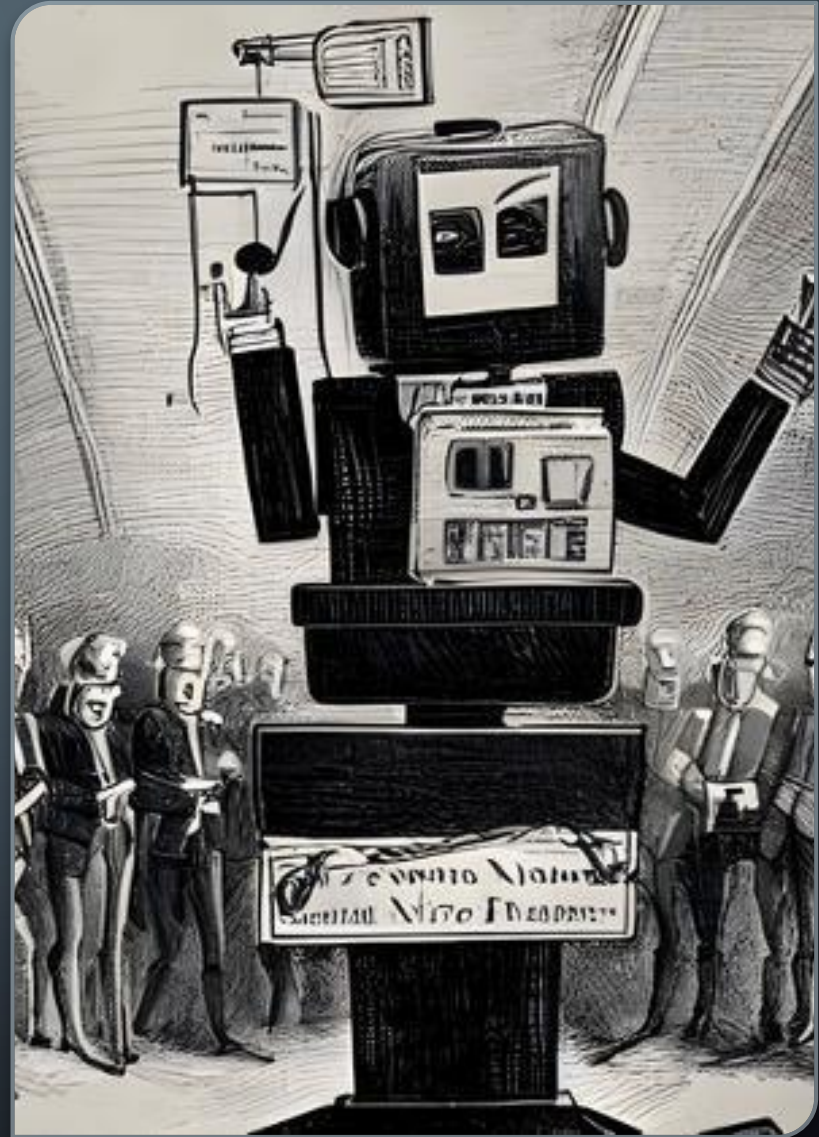# EXPLAINABLE AI & FAIRNESS

Thomas Souverain

7/12/2023

ROUNDT[AI]BLE

FORTN[AI]GHTLY NEWS

1. AI ACT

2. French scandal on social benefits

3. Google GEMINI

4. The Q* saga continues

5. Apple's MLX

2. FRENCH SCANDAL

ENQUÊTE | Depuis 2010, la CAF note ses treize millions de foyers d'allocataires pour orienter ses contrôles. « Le Monde » a enquêté sur ce système, accusé de discriminer les plus vulnérables.

Dis-moi qui tu es, l'algorithme dira si tu es suspect. A la Caisse nationale des allocations familiales (CNAF), où la recherche des erreurs de déclaration et des fraudes s'est industrialisée au cours des dernières années, un outil a été érigé en totem : le *data mining* (exploration de données). La priorisation des dossiers à contrôler repose aujourd'hui quasi exclusivement sur un « score de risque » calculé pour chaque allocataire suivant une batterie de critères personnels.

Ce système fait la fierté de l'institution, qui en vante les performances. Mais les alertes se multiplient sur les possibles dérives de cet algorithme, baptisé « data mining données entrantes » (DMDE). Plusieurs associations ont accusé la CNAF de discriminer les plus vulnérables au sein des publics à qui elle verse chaque mois le revenu de solidarité active (RSA), les aides au logement ou encore les allocations familiales.

L'ancien Défenseur des droits Jacques Toubon a fustigé en 2020 une démarche fondée sur des *« préjugés et des stéréotypes »*, tandis que de nombreux médias, du *Monde* à *Radio France*, en passant par *StreetPress*, ont documenté la détresse des allocataires face à un système implacable.

https://www.lemonde.fr/les-decodeurs/article/2023/12/04/profilage-et-discriminations-enquete-sur-les-derives-de-l-algorithme-des-caisses-d-allocations-familiales_6203796_4355770.html

https://deepmind.google/technologies/gemini/#introduction

Multimodal: text, code, audio, image and video.

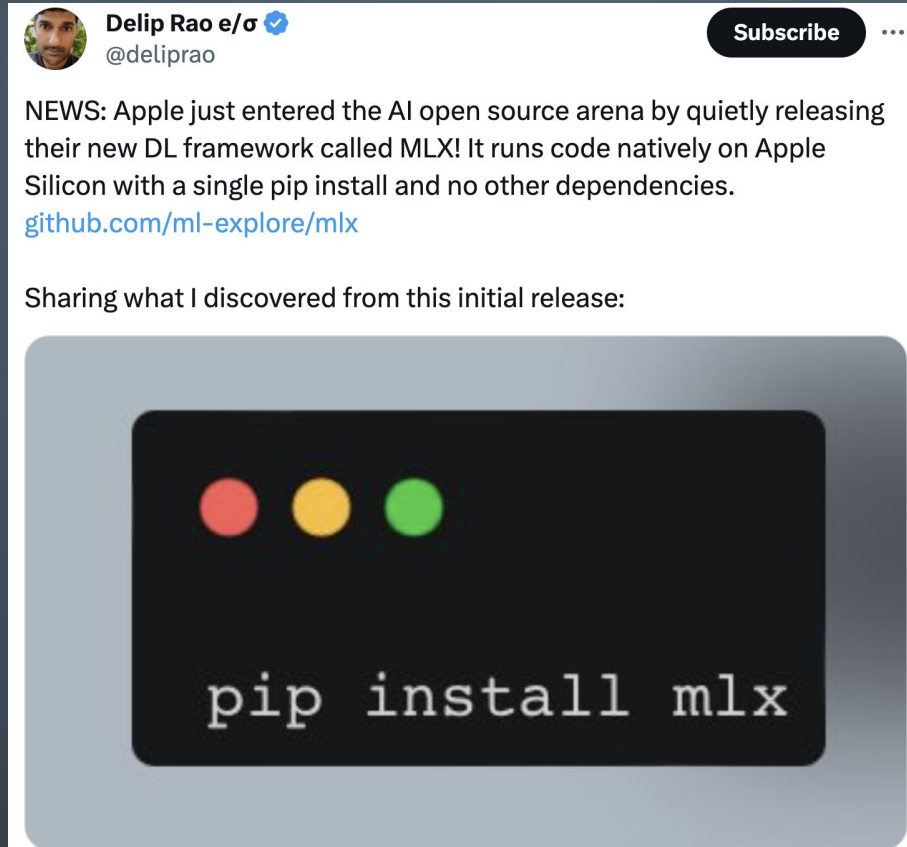**Models**: Ultra, Pro (GPT 3.5), Nano 1 & 2 (2 mobile versions)

Ultra Benchmarks:

- Outperforms GPT4 on most tasks
- Beats human experts on MMLU (Massive Multitask Language Understanding)
- Outperforms various models on multimodal tasks

Jürgen Schmidhuber ✓ @SchmidhuberAI · 2h ···

So @ylecun: "I've been advocating for deep learning architecture capable of planning since 2016" vs me: "I've been publishing deep learning architectures capable of planning since 1990." I guess in 2016 @ylecun also picked up the torch. (References attached)...

Show more

Yann LeCun ✓ ∞ ···
@ylecun

Please ignore the deluge of complete nonsense about Q*.
One of the main challenges to improve LLM reliability is to replace Auto-Regressive token prediction with planning.

Pretty much every top lab (FAIR, DeepMind, OpenAI etc) is working on that and some have already published ideas and results.

It is likely that Q* is OpenAI attempts at planning. They pretty much hired Noam Brown (of Libratus/poker and Cicero/Diplomacy fame) to work on that.

[Note: I've been advocating for deep learning architecture capable of planning since 2016].

10:02 PM · Nov 24, 2023 · **1.4M** Views

💬 257   ⟲ 930   ♡ 6.1K   🔖 1.9K   ⬆

💬 36   ⟲ 70   ♡ 387   � 41K   🔖   ⬆

Delip Rao e/σ ✓
@deliprao

Subscribe ···

NEWS: Apple just entered the AI open source arena by quietly releasing their new DL framework called MLX! It runs code natively on Apple Silicon with a single pip install and no other dependencies.
github.com/ml-explore/mlx

Sharing what I discovered from this initial release:

pip install mlx

Examples

The MLX examples repo has a variety of examples, including:

- Transformer language model training.
- Large-scale text generation with LLaMA and finetuning with LoRA.
- Generating images with Stable Diffusion.
- Speech recognition with OpenAI's Whisper.

[AI]NNOUNCEMENTS

Special Announcements from AI-PHI Community

- **Awareness in AI (AWAI) a special session of ICAART conference**
  Rome IT,
  24-26 February 2024

- Creation of a **new French Institute on AI for Society**

- **Conference on AI Regulations** next week at ENS

- Conference on the **Impact of Artificial Consciousness** in May 2024 at ENS

AWARENESS IN AI (AWAI) A SPECIAL SESSION OF ICAART CONFERENCE

Rome IT,
24-26 February 2024

The aim is to gather people from the same EU call "Awareness Inside" (such as VALAWAI, ASTOUND, etc.) and engage discussions with anyone who is interested on Awareness in AI (implementations, mechanisms, measures, applications, relations with theories of consciousness)

Submissions are due by the 21st of December

https://icaart.scitevents.org/AWAI.aspx

MARDI DU GRAND CONTINENT · 05 DÉCEMBRE 2023

La société face aux défis de l'IA

Jamal Atif · Anne Bouverot · Yann Le Cun · Marc Mézard ·
Gabriela Ramos · Martin Tisné · Frédéric Worms

Fondation Abeonă
Championing Responsible AI

ENS | PSL★

Dauphine | PSL★

NEW INSTITUTE ON AI FOR SOCIETY

CONFERENCE ANNOUNCEMENT

AI & SOCIETY INITIATIVE

MENU ≡   FRANÇAIS

Debate on AI Regulations

The extremely rapid development of Artificial Intelligence in recent months is both astonishing and concerning. It seems important to regulate this field to avoid potential abuses. But what exactly should be regulated? While almost everyone agrees on the need to strengthen the regulation of applications that incorporate AI techniques, the regulation of foundational models (particularly large language models) remains much more debated. These models are not applications per se; they are complex but can lead to a wide variety of services in different fields (law, healthcare, education, etc.). Should they be regulated? If so, which aspects of these models should be regulated? Or should all legislation focus on the applications, leaving the models without specific rules? More broadly, could regulation risk stifling innovation, as is often claimed?

Participants

- Raja Chatila (Président du comité national consultatif d'éthique du numérique)
- Camilla Penzo (PEReN, Pôle de'Expertise sur la Régulation du Numérique)
- Isabelle Ryl (Directrice de Prairie)
- Anastasia Stasenko (Research Lead à opsci.ai)

Practical details

What ? Discussion - Debate

Where ? Ecole normale supérieure, Amphithéâtre Jaurès, Département d'Etudes Cognitives, 29 rue d'Ulm 75005 Paris

When ? Thursday, December 14th, from 6:00 PM to 7:30 PM

Registration : Free but mandatory here \

The debates will be transmitted in a hybrid format if conditions allow. Participation in person is recommended.

https://ais-initiative.github.io/2023-ai-regulation

CONFERENCE ANNOUNCEMENT

May 29th-31st : Impact of Artificial Consciousness on Society

June 1st: Artificial Consciousness for the General Public (French)